



COLEGIO
LIBRE DE
EMERTOS

TRES DÉCADAS DE TRANSGÉNICOS

Director: Francisco García Olmedo



EL GENOMA VEGETAL

Y SU EXPRESIÓN

(Conferencia II)

Francisco García Olmedo
Real Academia de Ingeniería

Introducción

Hasta principios de este siglo, la biología molecular de las plantas se había centrado en el estudio de genes individuales que codifican proteínas más o menos relevantes o que determinaban características biológicas de interés. Con la publicación en diciembre del año 2000 de la secuencia completa del genoma de la crucífera modelo *Arabidopsis thaliana*, la escala de observación pasó bruscamente del gen aislado al genoma completo. Los avances recientes en la investigación del genoma vegetal, de la secuencia del ADN que lo compone, de su estructura y de sus funciones han revolucionado el conocimiento botánico y el progreso tecnológico: elucidación de relaciones evolutivas y filogenéticas, descubrimiento de nuevos genes y alelos génicos, genotipados mediante SNP (Single Nucleotide Polymorphism; polimorfismos de nucleótido único) y, en el aspecto práctico, la mejora vegetal molecular.

A la secuencia del genoma de la mencionada crucífera pronto siguieron las de otras especies: el arroz en 2005, el chopo en 2006, la vid en 2007, o la papaya en 2008. En el mismo 2008, un número monográfico de la revista *International Journal of Plant Genomics* daba noticias de más de una docena de genomas correspondientes a especies vegetales de todo tipo y en la actualidad el ritmo de aparición de información genómica es vertiginoso, gracias sobre todo al extraordinario abaratamiento y a la extrema simplificación de las técnicas de secuenciación del ADN.

En esta lección nos proponemos describir de un modo simplificado en qué han consistido estos avances básicos. En concreto, nos referiremos brevemente al gen

como unidad funcional fundamental, luego describiremos la estructura del genoma vegetal y finalmente explicaremos cómo extraer significado biológico y aplicaciones prácticas a partir de la información genómica.

Genes

Como se ha repetido, la información genética está cifrada en la sucesión con de los cuatro distintos eslabones eslabones (las bases A,T,G,C) que componen una cadena de ADN. Hace ya casi medio siglo que se descifró la clave genética que rige el flujo de información del ADN, compuesto de bases, a la proteína, compuesta de aminoácidos, mediado por la transcripción a ARN, cuyo lenguaje es una mero dialecto de aquel del ADN. Un gen es, por tanto, un tramo de ADN que en general determina una proteína con una secuencia de aminoácidos determinada, de acuerdo con las correspondencias plasmadas en la clave genética. No todo el tramo de ADN al que llamamos gen en términos moleculares se transcribe y se traduce; sólo lo hace una región que denominamos 'codificante'. Esta región está precedida por otra que se designa 'promotor', y seguida por un tramo donde residen las señales de terminación del mensaje. En la región codificante, que en organismos superiores puede ser discontinua, reside la información que determina la estructura del producto génico (proteína), mientras que en el promotor se agrupa la mayor parte de los elementos 'informáticos' del gen, los cuales determinan en qué momento del desarrollo del organismo, o en respuesta a qué estímulo externo, en qué cantidad y en qué células se ha de fabricar la correspondiente proteína codificada por él. Los agentes que reconocen estas instrucciones son a su vez unas proteínas codificadas por otros genes que se llaman factores de transcripción y que regulan la actividad de la maquinaria de transcripción. Los factores de transcripción reconocen secuencias de

bases concretas dentro de los promotores correspondientes y realizan su función uniéndose a ellas.

De lo dicho en el párrafo anterior se deduce que hay genes, a los que llamaremos reguladores, que determinan la activación o la inhibición de otros genes. De hecho, existen redes jerárquicas de poderes y contrapoderes, redes informáticas de gran complejidad que incluyen también otros tipos de proteínas reguladoras. Éstas pueden dar cuenta del funcionamiento de un ser vivo complejo.

La proteína codificada por un gen regulador, lo que hemos llamado un factor transcripcional, puede tener un efecto positivo sobre la actividad de un gen diana, estimulando la producción de la proteína correspondiente, y negativo sobre otro gen diana, inhibiendo su expresión. Un factor de transcripción puede necesitar para ejercer su función la ayuda de otra proteína o de un ión que actúe de señal. En este caso el mecanismo es el de un interruptor que se puede accionar por señales externas o internas.

El genoma vegetal

Las plantas no solo poseen un genoma nuclear, que es el más grande y complejo, sino que los orgánulos subcelulares vegetales, denominados cloroplastos y mitocondrias, también poseen sus respectivos genomas. Un genoma dado puede caracterizarse por su tamaño, que es muy variable entre los genomas nucleares vegetales, su contenido génico, la naturaleza y disposición de sus secuencias repetidas y la presencia de rastros de antiguas duplicaciones y poliploidizaciones (multiplicación de un genoma básico y agregación de genomas estrechamente

relacionados). La enorme variabilidad del tamaño de los genomas nucleares vegetales se debe en gran medida a la mayor o menor abundancia de secuencias repetidas y de retrotransposones, secuencias parecidas a las de los retrovirus que tienen gran facilidad para cambiar de lugar en el genoma. Esta variación en el contenido genómico de secuencias repetidas no tiene un correlato funcional obvio, es decir, no se sabe bien qué consecuencias tiene para la fisiología de la especie. El contenido de estas secuencias puede ser relativamente bajo en una especie como *Arabidopsis thaliana* y, en cambio, ser muy alto en el trigo (Tabla 1).

Tabla 1. Tamaño del genoma nuclear en distintas especies vegetales

Nombre común	Nombre científico	Tamaño del genoma Megabases (Mb, 10 ⁶ bases)
Trigo	<i>Triticum aestivum</i>	15.966
Maíz	<i>Zea mays</i>	2.292
Tomate	<i>Solanum lycopersicum</i>	907
Manzana	<i>Malus X domestica</i>	743
Vid	<i>Vitis vinífera</i>	483
Arabidopsis	<i>Arabidopsis thaliana</i>	120
Ser humano	<i>Homo sapiens</i>	2.910

Cuatro de cada cinco especies vegetales con flor ha sufrido un proceso de poliploidización en el curso de su evolución, ya sea por duplicación de su genoma completo o por agregación de genomas próximos entre sí (aloploidización). Este es el caso de especies como el maíz, la patata o el trigo, que se han poliploidizado más o menos recientemente en el tiempo evolutivo. El trigo panificable es un buen ejemplo de aloploidización repetida y reciente: a partir de dos especies de 7 pares de cromosomas se formó el trigo alotetraploide, *Triticum durum*, usado en la elaboración de pastas alimenticias, y por la adición de otro genoma de 7 pares de cromosomas se formó el alohexaploide *T. aestivum*, con 21 pares de cromosomas,

cuyo uso principal es la fabricación de pan. La redundancia génica que han supuesto todos estos procesos de duplicación ha permitido la creación de gran variabilidad genética y de capacidad de adaptación evolutiva en las especies que la han sufrido.

Aunque en la actualidad se están secuenciando los genomas de todas las especies vegetales de interés económico, el desarrollo de la genómica ha gravitado y seguirá gravitando sobre especies modelo, como las silvestres *Arabidopsis thaliana* o *Brachypodium distachyon*, y especies gran cultivo como la cebada, *Hordeum vulgare*, o el arroz, *Oryza sativa*. La dicotiledónea *Arabidopsis* reúne una serie de propiedades que la hacen ideal como especie modelo: genoma pequeño (poco ADN repetitivo; genes próximos entre sí), ciclo de 6 semanas de semilla a semilla, y sólo 5 cromosomas. Del arroz existen dos subespecies, la *indica* y la *japonica*. El arroz tiene un genoma más pequeño que el de la cebada, 420 frente a 5.000 megabases, y ambas especies presentan menos ventajas para la secuenciación que *Arabidopsis*, pero son buenos representantes de las monocotiledóneas y están entre las principales cosechas mundiales. Ambas especies tienen en común que fáciles de transformar y que se prestan a todo el amplio repertorio de técnicas encaminadas a dilucidar el significado de las distintas secuencias génicas.

Mapas y secuencias de ADN

Pronto en la historia de la genética mendeliana se desarrollaron métodos para obtener mapas genéticos, es decir, para establecer el orden que ocupan los genes en los cromosomas y las distancias relativas entre unos y otros. Con el desarrollo de técnicas para clonar tramos anónimos del ADN de un genoma se abrió la posibilidad de aplicar el método mendeliano de mapeo para ordenar estos tramos anónimos de

ADN y obtener así mapas físicos del genoma con una alta densidad de marcadores. Si en la descendencia de un cruzamiento dado un carácter macroscópico, como el color de un pétalo o la forma de una hoja, se hereda concomitantemente con una variante determinada de un marcador molecular, podremos seguirle la pista al carácter macroscópico de interés siguiéndosela al marcador apropiado, lo que para ciertos caracteres, como pueda ser la resistencia a una enfermedad o el rendimiento de grano, resulta una forma muy eficaz de manejarlos en la práctica. La mejora vegetal clásica asistida por marcadores es una práctica habitual en nuestros días, especialmente en lo que se refiere a caracteres cuantitativos, como la altura de la planta, el tamaño del grano o el rendimiento de la cosecha. El carácter de interés no está necesariamente codificado por el tramo de ADN que lo marca, pero el gen responsable se tiene que encontrar necesariamente en la proximidad inmediata del lugar del mapa correspondiente al marcador. Esto hace que, además de ayudarnos al manejo empírico del carácter en cuestión, el marcador nos puede servir de punto de partida para buscar y clonar el gen de interés.

La secuenciación de genomas completos ha avanzado extraordinariamente gracias al desarrollo de técnicas clave: el uso de marcadores fluorescentes, la robotización de las manipulaciones necesarias y la implementación del software y el hardware necesarios para el manejo y procesamiento de cantidades ingentes (gigabytes) de datos puntuales. No es este el lugar para referirnos a los fundamentos técnicos del vertiginoso avance de las técnicas de secuenciación; baste señalar que la primera secuenciación del genoma humano en 2003 costó por encima de los 500 millones de dólares mientras que ahora estamos en la era del 'genoma a mil dólares'. Los algoritmos informáticos desarrollados permiten estimar el número de genes

presumiblemente contenidos en un genoma dado, pero eventualmente será necesario refinar estos resultados y caracterizar la acción génica por la vía experimental de la genómica funcional, lo que abordaremos más adelante.

El orden de los genes a lo largo de los cromosomas es una característica bastante conservada durante la evolución, de modo que las especies correspondientes a un grupo taxonómico tienden a tener genes equivalentes en posiciones equivalentes de los cromosomas. A esta característica se la denomina 'colinearidad'. La secuenciación de los genomas de especies próximas entre sí ha permitido establecer relaciones de colinearidad dentro de distintos grupos de especies: entre los cereales (trigo, arroz, cebada y maíz), entre las leguminosas (judías, guisante, soja), o entre las crucíferas (col, colza, brócoli, Arabidopsis). La colinearidad facilita el manejo básico y aplicado de la información genómica.

Genómica funcional

Todas las células de un organismo dado tienen esencialmente el mismo genoma, los mismos genes, pero unas células y otras expresan distintos subconjuntos del acervo génico de la especie. Ciertos genes se expresan en todo tipo de células porque atienden al repertorio básico de funciones necesarias para la integridad celular mientras que otros se expresan selectivamente en cada tipo de células para conferirle la funcionalidad que le es característica: en las células de un pétalo y en las de la hoja se expresan una serie de genes comunes y otros de forma específica, de modo que en las del pétalo se formen por ejemplo ciertos pigmentos y en las de la hoja no lo hagan. Una serie de técnicas permiten estudiar cualitativa y cuantitativamente la expresión génica en los distintos tipos celulares.

Se ha indicado que en la expresión génica la información codificada en un gen (en el ADN del genoma) se 'transcribe' a un ARN mensajero que a su vez será 'traducido' a la proteína correspondiente. El conjunto de las proteínas existente en una célula en un momento dado y sus interacciones determinan el 'metabolismo' celular. La genómica funcional se ocupa de estudiar el conjunto de la expresión génica en distintos tipos de células en las distintas fases de dicha expresión: la 'transcriptómica', el estudio del RNA transcrito mediante la técnica de las 'etiquetas de secuencias expresadas' (ESTs; *Expressed Sequence Tags*); la 'proteómica', mediante la separación, caracterización y cuantificación del acervo proteico de un tipo celular en un momento dado y la 'metabolómica', cuando se investigan los metabolitos y las cantidades que de ellos están presentes en un tipo celular. El conjunto de las 'ómicas', como se las suele llamar, permite extraer significado biológico de los crípticos datos de secuencia.

El conjunto de las ESTs obtenidos de un tipo celular es como una 'fotografía' de la actividad génica en el momento de observación: se obtienen etiquetas de los RNAs correspondientes a los genes expresados en ese momento. El examen de estas etiquetas permite establecer relaciones de homología (parecido) con genes caracterizados en otras especies y obtener pistas sobre la función biológica del gen en cuestión. Son también herramientas útiles en el descubrimiento de nuevos genes, aunque la técnica es poco sensible para detectar genes que se expresan a un nivel bajo. Las ESTs permiten adicionalmente el establecimiento de mapas de expresión/transcripción que dan información sobre cómo se distribuyen en el genoma los genes que codifican las enzimas de una misma ruta metabólica.

Se han encontrado ESTs para aproximadamente el 60 % de los genes de Arabidopsis, lo que sirve de base para la obtención de los llamados 'perfiles transcripcionales'. Estos perfiles suponen el análisis en paralelo de la expresión de miles de genes en respuesta a situaciones biológicas distintas, desvelando procesos y rutas metabólicas hasta ahora desconocidas. Es decir, mediante un repertorio técnico que no describiremos aquí, las ESTs permiten analizar cuáles entre los miles de genes contenidos en el genoma de una planta se expresan de un modo específico en situaciones tales como la infección por un patógeno determinado, la exposición al frío o al calor, o durante el inicio de la formación de los primordios florales a partir del meristemo.

La aproximación mendeliana a la genética partía de un carácter dado y trataba de descifrar su base genética. La nueva tecnología, en cambio, permite invertir el proceso y practicar lo que se denomina 'genética reversa', lo que consiste en determinar la función de un gen para el que se conoce la secuencia. En general, existen dos formas de investigar la función de un gen: impidiendo su función en su contexto habitual de expresión o expresando el gen fuera de contexto en un tejido o en una especie donde habitualmente no se expresa. Por distintos trucos que aquí no explicaremos es posible inutilizar un gen de modo que no se produzca la proteína correspondiente y entonces podemos preguntarnos qué alteraciones en el normal desarrollo y funcionamiento de la planta se han producido por el fallo del gen o, alternativamente, es también posible expresar un determinado gen fuera de lugar, en un tipo celular donde normalmente no se exprese, y observar qué consecuencias tiene esa expresión exógena para el normal funcionamiento de esas células.

Hay tres factores principales que limitan los resultados obtenidos por genética inversa: a) la inutilización de ciertos genes esenciales para el funcionamiento celular resulta letal y no rinde información útil; b) muchos genes son funcionalmente redundantes, por lo que su inutilización no produce ningún fenotipo observable; y c) las mutaciones de ciertos genes tienen efectos múltiples que dificultan grandemente el análisis de la función básica de un gen dado.

La transgénesis y la genómica han abierto la posibilidad de diseccionar las complejas redes regulatorias que subyacen a los procesos de desarrollo de la planta y de las respuestas de ésta a los retos que plantea su relación con el entorno, sean retos bióticos o abióticos. De particular interés ha sido la investigación funcional de los distintos tipos de factores de transcripción y de su combinatoria en los más variados procesos. En la siguiente lección pasaremos revista a los avances más notables a este respecto.

Dos ejemplos ilustrativos

Para concretar algunos aspectos de la genómica vegetal, glosaremos los progresos principales respecto a los genomas de dos especies, una dicotiledónea silvestre, *Arabidopsis*, y una monocotiledónea cultivada, cebada:

Arabidopsis thaliana

Ya se ha dicho que el genoma del ecotipo Columbia de *Arabidopsis* fue el primero de origen vegetal en ser secuenciado por completo. El proyecto se inició en 1996 por un consorcio internacional y la secuencia fue publicada en el año 2000. Esta especie tiene 5 cromosomas cuya longitud varía entre 17.549.867 pares de bases

(cromosoma 4) y 29.105.111 pares de bases (cromosoma 1). El número total de genes se estima en unos 25.000 que se distribuyen de un modo irregular por los cinco cromosomas con una densidad media que oscila entre 4.000 y 4.500 pares de bases por gen. Esta densidad es mucho más alta que la de la cebada, especie en la que los genes están más espaciados entre sí. Así se explica que dos genomas de tamaño tan distinto den lugar a especies de un número de genes y una complejidad funcional del mismo orden.

El genoma de *Arabidopsis* ha servido de banco de pruebas para el desarrollo de todas las ómicas y para el esclarecimiento de buena parte de los principales mecanismos que rigen la vida de las plantas. En la siguiente lección pasaremos revista a estos mecanismos.

Hordeum vulgare

Los cereales siguen siendo en la actualidad el grupo de cosechas más importante para nuestro sustento, con una producción superior a los 2.000 millones de toneladas en 2006. La cebada fue domesticada en el Creciente Fértil hace unos diez milenios. Desde su domesticación, ha venido sirviendo como pienso animal, como ingrediente básico en la fabricación de cerveza y como alimento humano, aunque este último uso haya declinado en tiempos recientes. A lo largo de milenios, los agricultores fueron adaptando las poblaciones locales a sus necesidades, dando lugar a un alto número de variedades autóctonas. A principios del siglo XX se usaron estas variedades primitivas para someterlas a una mejora sistemática mediante cruzamiento y selección que han ido multiplicando su rendimiento de una forma

paulatina. Este progreso empírico puede y debe reforzarse gracias aun conocimiento más completo de la genética de la especie.

La genómica de la cebada es un buen exponente de los avances en la biología y la mejora molecular de este tipo de cosechas. Entre los progresos realizados a este respecto pueden citarse los siguientes: a) una acumulación considerable de datos sobre sus ESTs; b) extensos estudios de transcriptómica, proteómica y metabolómica; c) desarrollo de nuevas técnicas de modelaje; d) técnicas de transformación eficientes y disponibilidad y colecciones de líneas con genes inutilizados (*knock outs*) que cubren buena parte del genoma; e) La secuencia completa del genoma.

En el futuro pueden esperarse nuevos progresos basados en la aplicación de los hallazgos de la 'biología de sistemas' a la mejora molecular de esta especie y de sus parientes próximos

Bibliografía

- PK Gupta, Y. Xu (2008) *Genomics of major crops and model plant species*. International Journal of Plant Genomics 2008:1-2
- T. Ferrier et al. (2011) *Arabidopsis paves the way: genomic and network analyses in crops*. Current opinion in Biotechnology 22:260-270
- N Sreenivasulu et al. (2008) *Barley genomics: An overview*. International Journal of Planta Genomics 2008.1-15